Low-Signal-Energy Asymptotics of Capacity and Mutual Information for the Discrete-Time Poisson Channel

Alfonso Martinez

Abstract

The first terms of the low-signal-energy asymptotics for the mutual information in the discrete-time Poisson channel are derived and compared to an asymptotic expression of the capacity. In the presence of non-zero additive noise (either Poisson or geometric), the mutual information is concave at zero signal-energy and the minimum energy per bit is not attained at zero capacity. Fixed signal constellations which scale with the signal energy do not attain the minimum energy per bit. The minimum energy per bit is zero when additive Poisson noise is present and $\varepsilon_n \log 2$ when additive geometric noise of mean ε_n is present.

I. MOTIVATION AND NOTATION

In the complex-valued Gaussian channel with signal-to-noise ratio SNR the mutual information of very general constellations (e. g. zero-mean with uncorrelated real and imaginary parts each of energy $\frac{1}{2}$ [1]) has the same low-SNR asymptotics as the channel capacity, namely $\mathrm{SNR} - \frac{1}{2}\mathrm{SNR}^2 + \mathrm{o}(\mathrm{SNR}^2)$. These constellations also attain the minimum bit-energy-to-noise-variance ratio of -1.59 dB at vanishing SNR. A natural question concerns the extent to which this universality extends to other common channel models. We consider here the discrete-time Poisson channel, frequently used to represent optical communication channels, and quantity the gap between the channel capacity and the mutual information for fixed signal constellations. As a by-product of our analysis, we also determine the asymptotic form of the capacity at vanishing signal energy.

Consider a memoryless channel with input X and output Y given by the sum

$$Y = S(X) + Z \tag{1}$$

of a noise Z and a signal component $S(X)$, itself a function of the input X. The input X is a non-negative real number (i. e. it has units of energy), drawn from a unit-energy set $\mathcal{X}$ according to a probability distribution $P(x)$. We let $S(X)$ be distributed according to a Poisson distribution of parameter $\varepsilon_s X$, where ε_s is an average signal energy. The output components $S(X)$, Z, and Y are nonnegative integers.

A. Martinez is with Centrum Wiskunde & Informatica, The Netherlands. e-mail: alfonso.martinez@ieee.org.

We study three channel models: noiseless, with $Z = 0$; additive Poisson noise, where Z follows a Poisson distribution of mean $\varepsilon_n > 0$; and additive geometric noise, with Z distributed according to a geometric distribution of mean $\varepsilon_n > 0$. With additive Poisson noise the channel transition probability, denoted by $Q(y|x)$, is given by

$$Q(y|x) = e^{-(\varepsilon_s x + \varepsilon_n)} \frac{(\varepsilon_s x + \varepsilon_n)^y}{y!}, \tag{2}$$

where $\varepsilon_n \geq 0$. For the channel with geometric noise, we have

$$Q(y|x) = \sum_{l=0}^{y} \frac{e^{-x}}{1 + \varepsilon_n} \left(\frac{\varepsilon_n}{1 + \varepsilon_n} \right)^y \frac{\left(x\left(1 + \frac{1}{\varepsilon_n}\right) \right)^l}{l!}. \tag{3}$$

Remark that the model with additive geometric noise arises in representations of electromagnetic radiation as a photon gas [2].

In this letter, we compute the minimum energy per bit for these models. We also study the asymptotics of the mutual information $I\big(X; S(X) + Z\big)$ at low ε_s and compare them with the channel capacity $\mathrm{C}(\varepsilon_s)$ at energy ε_s. The main results are presented in the next section; the proofs can be found in the appendices.

II. MAIN RESULTS AND DISCUSSION

A. Capacity Asymptotics and Capacity per Unit Cost

A closed-form expression for the capacity $\mathrm{C}(\varepsilon_s)$ of the discrete-time Poisson channel is not known. For the noiseless channel (i. e. with $Z = 0$), the best known firm upper bound was derived in [3], and is given by

$$\mathrm{C}(\varepsilon_s) \leq \log\left(\left(1 + \frac{\sqrt{2e} - 1}{\sqrt{1 + 2\varepsilon_s}} \right) \frac{\left(\varepsilon_s + \frac{1}{2}\right)^{\varepsilon_s + \frac{1}{2}}}{\sqrt{e}\varepsilon_s^{\varepsilon_s}} \right). \tag{4}$$

As for lower bounds, binary modulation attains a high mutual information at low values of ε_s. Specifically, let the symbols be located at $x = 0$ and $x = 1/p$, and be respectively used with probabilities $1 - p$ and p. We denote the mutual information attained by such modulation by $I_\mathrm{b}(p)$. A trite computation gives

$$I_\mathrm{b}(p) = -(p - pe^{-\frac{\varepsilon_s}{p}}) \log p - \varepsilon_s e^{-\frac{\varepsilon_s}{p}} - (1 - p + pe^{-\frac{\varepsilon_s}{p}}) \log(1 - p + pe^{-\frac{\varepsilon_s}{p}}). \tag{5}$$

Setting $p = \varepsilon_s$, and using the Taylor expansion of the logarithm $\log(1 + x)$ around $x = 0$ in both Eq. (4) and Eq. (5), we have that

$$-(1 - e^{-1})\varepsilon_s \log \varepsilon_s + \mathrm{O}(\varepsilon_s) \lesssim \mathrm{C}(\varepsilon_s) \lesssim -\varepsilon_s \log \varepsilon_s + \mathrm{O}(\varepsilon_s). \tag{6}$$

Therefore, the capacity $\mathrm{C}(\varepsilon_s)$ of the noiseless discrete-time Poisson channel behaves as $\mathrm{O}(-\varepsilon_s \log \varepsilon_s)$ at vanishing ε_s.

Moreover, a similar reasoning shows that flash signalling with $p = -\varepsilon_s \log \varepsilon_s$ (for $\varepsilon_s \ll 1$) asymptotically behaves as

$$I_\mathrm{b}(-\varepsilon_s \log \varepsilon_s) = -\varepsilon_s \log \varepsilon_s + \mathrm{o}(-\varepsilon_s \log \varepsilon_s) \tag{7}$$

for low ε_s. Combining Eqs. (6) and (7), we obtain the following

Theorem 1. *For vanishing ε_s the capacity $C(\varepsilon_s)$ behaves as*

$$C(\varepsilon_s) = -\varepsilon_s \log \varepsilon_s + o(-\varepsilon_s \log \varepsilon_s). \tag{8}$$

This result complements the asymptotic behaviour for very large values of ε_s, which was established in [4], [3],

$$C(\varepsilon_s) = \frac{1}{2} \log \varepsilon_s + o(\log \varepsilon_s). \tag{9}$$

Observe that Eq. (6) implies that the capacity per unit energy C_1,

$$C_1 = \sup_{\varepsilon_s} \frac{C(\varepsilon_s)}{\varepsilon_s}. \tag{10}$$

is infinite for the discrete-time Poisson channel. This well-known result had been obtained by Verdú [5] by exploiting a simple formula for C_1 in channels which have a zero-energy symbol ($x = 0$ in our case), namely

$$C_1 = \sup_x \frac{D\big(Q(y|x)||Q(y|x = 0)\big)}{\varepsilon_s x}, \tag{11}$$

where $D\big(Q(y|x)||Q(y|0)\big)$ is the divergence between the transition probabilities $Q(y|x)$ for arbitrary input x and zero input $x = 0$.

By definition, the minimum energy per bit $\varepsilon_{b,\min} = \inf_{\varepsilon_s} \frac{\varepsilon_s}{C(\varepsilon_s)}$, where the capacity is measured in bits, is given by $\varepsilon_{b,\min} = \frac{\log 2}{C_1}$.

Applied to the family of channels we consider, we have

Theorem 2. *In the absence of additive noise, i. e. for $Z = 0$, or in the presence of additive Poisson noise, the minimum energy per bit is $\varepsilon_{b,min} = 0$. Equivalently the capacity per unit cost is $C_1 = \infty$.*

With additive geometric noise, the minimum energy per bit is $\varepsilon_{b,min} = \varepsilon_n \log 2$ and the capacity per unit cost is $C_1 = \varepsilon_n^{-1}$ nats.

The latter result is new. Remarkably, the minimum energy per bit has the same form as in the Gaussian channel, for a minimum ratio $\varepsilon_b/\varepsilon_n$ of -1.59 dB.

Proof: The proof can be found in Appendix A. ∎

B. Mutual Information Asymptotics

We now move on to study the asymptotics of the mutual information $I\big(X; S(X) + Z\big)$ (in nats) at low ε_s for fixed unit-energy constellations $\mathcal{X}$. The output $S(X)$ is distributed according to a Poisson distribution of parameter $\varepsilon_s X$. We determine the first two coefficients c_1 and c_2 in the Taylor expansion around $\varepsilon_s = 0$, that is

$$I\big(X; S(X) + Z\big) = c_1 \varepsilon_s + c_2 \varepsilon_s^2 + o(\varepsilon_s^2). \tag{12}$$

Also, we define the energy per bit ε_b as $\varepsilon_b = \frac{\varepsilon_s}{I(X;Y)} \log 2$.

Denoting the first- and second-order moments of the constellation by μ_1 (we often have $\mu_1 = 1$) and μ_2 respectively, we have

Theorem 3. *In the absence of additive noise, i. e. for $Z = 0$, the mutual information behaves at low ε_s as Eq. (12)*

with

$$c_1 = \sum_{x \in \mathcal{X}} P(x) x \log \frac{x}{\mu_1}, \quad c_2 = \frac{1}{2} \left(\mu_2 - \mu_1^2 - \mu_2 \log \frac{\mu_2}{\mu_1^2} \right). \tag{13}$$

As long as $\varepsilon_n > 0$, regardless of whether Z has a Poisson or a geometric distribution, the coefficients in Eq. (12)

are

$$c_1 = 0, \quad c_2 = \frac{1}{2} \frac{(\mu_2 - \mu_1^2)}{\varepsilon_n}. \tag{14}$$

Proof: Eq. (13) is proved in Appendix B, Eq. (14) in Appendices C and D for Poisson and geometric noise respectively. ∎

C. Discussion

In the presence of non-zero additive noise, the mutual information is concave at zero signal-energy (because $c_1 = 0$, $c_2 > 0$) and the minimum energy per bit is not attained at zero capacity. This effect can be seen in Fig. 1, which depicts the energy per bit ε_b as a function of the mutual information for several positive values of ε_n. The modulation depicted is uniform pulse-energy modulation (PEM), e. g. 2^m points uniformly located between 0 and 2 with spacing $1/2^m$. General signal constellations do not attain the minimum energy-per-bit at vanishing signal energy. The determination of the minimum energy per bit attained by these modulations is an open problem.

[Fig. 1 about here.]

Moreover, since there is no coefficient in $-\varepsilon_s \log \varepsilon_s$ in Eq. (12), these modulations do not attain the minimum energy per bit $\varepsilon_{b,\min}$. This is true even for the noiseless channel, for which $c_1 \neq 0$. In this case, binary modulation at points $x = 0$ and $x = 1/p$ respectively used with probabilities $1 - p$ and p, with fixed p, has coefficients c_1 and c_2

$$c_1 = -\log(p), \quad c_2 = \frac{1 - p + \log(p)}{2p}. \tag{15}$$

In the limit $p \to 0$, $c_1 \to \infty$ and the bit energy at zero capacity, $\varepsilon_{b,0}$, approaches 0, the minimum energy per bit. Fig. 2 depicts ε_b as a function of the mutual information for various fixed values of p and for $p = -\varepsilon_s \log \varepsilon_s$. For comparison, we also include the value of ε_b corresponding to the upper bound in Eq. (4). Even though $\varepsilon_{b,0}$ indeed approaches zero, it does so rather slowly. Also, the gap between the energy per bit corresponding to $I_b(-\varepsilon_s \log \varepsilon_s)$ and $\varepsilon_{b,\min}$ closes slowly. Numerical evaluation shows that it is only for values of ε_s below 10^{-307} (!) that I_b exceeds $0.99 \cdot \varepsilon_s \log \varepsilon_s^{-1}$. Even though one eventually has $C(\varepsilon_s) \simeq -\varepsilon_s \log \varepsilon_s$, convergence to the limit is very slow. This fact, together with the concave nature of the mutual information $I(X; Y)$ at zero ε_s for nonzero additive noise, suggests that the asymptotic analysis of the capacity and the mutual information in the discrete-time Poisson channel fails to capture the key features of these quantities. This behaviour stands in contrast with the Gaussian channel, where asymptotic expansions give an accurate representation of the capacity and the mutual information [6].

[Fig. 2 about here.]

APPENDIX A

CAPACITY PER UNIT ENERGY

We first consider the case with additive Poisson noise. Using Eq. (2) for $Q(\cdot|\cdot)$ and the definition of divergence, we have

$$D\big(Q(y|x)||Q(y|0)\big) = -\varepsilon_s x + (\varepsilon_s x + \varepsilon_n) \log \frac{\varepsilon_s x + \varepsilon_n}{\varepsilon_n}. \tag{16}$$

Hence,

$$\mathrm{C}_1 = \sup_x \left(-1 + \left(1 + \frac{\varepsilon_n}{\varepsilon_s x} \right) \log \frac{\varepsilon_s x + \varepsilon_n}{\varepsilon_n} \right) = \infty. \tag{17}$$

We now consider the channel with additive geometric noise. Using Eq. (3) for $Q(\cdot|\cdot)$ and the definition of divergence, we have

$$D\big(Q(y|x)||Q(y|0)\big) = \sum_y Q(y|x) \log \left(e^{\frac{\varepsilon_s x}{\varepsilon_n}} \sum_{l=0}^{y} e^{-\alpha} \frac{\alpha^l}{l!} \right), \tag{18}$$

where $\alpha = \varepsilon_s x \big(1 + \frac{1}{\varepsilon_n} \big)$. Let us define $P(l) = e^{-\alpha} \frac{\alpha^l}{l!}$ and the quantity $q(y) = \sum_{l=0}^{y} P(l)$, i. e. the cumulative distribution function of a Poisson random variable with mean α.

Moving the exponential out of the logarithm, we obtain

$$D\big(Q(y|x)||Q(y|0)\big) = \frac{\varepsilon_s x}{\varepsilon_n} + \sum_y Q(y|x) \log\big(q(y)\big). \tag{19}$$

Hence, the capacity per unit energy is given by

$$\mathrm{C}_1 = \frac{1}{\varepsilon_n} + \sup_x \frac{\sum_y Q(y|x) \log\big(q(y)\big)}{\varepsilon_s x}. \tag{20}$$

Since $q(y) \leq 1$, its logarithm is always non-positive, and

$$\mathrm{C}_1 \leq \frac{1}{\varepsilon_n}. \tag{21}$$

The proof is completed by proving that

$$\lim_{x \to \infty} \frac{e^{\frac{\varepsilon_s x}{\varepsilon_n}}}{\varepsilon_s x (1 + \varepsilon_n)} \left(\sum_y \left(\frac{\varepsilon_n}{1 + \varepsilon_n} \right)^y q(y) \log\big(q(y)\big) \right) = 0, \tag{22}$$

where we expressed $Q(y|x)$ as a function of $q(y)$. If this condition holds true, then Eq. (21) becomes an equality.

In Eq. (22) we split the summation over y into two parts, from 0 to $y^* = \lfloor \alpha \rfloor$, and from $y^* + 1$ to infinity. In the first part, $e^{-\alpha} \frac{\alpha^y}{y!}$ is an increasing function in y, and therefore

$$q(y) = \sum_{l=0}^{y} P(l) \geq \sum_{l=0}^{y} P(0) = (y + 1) P(0) = (y + 1) e^{-\alpha}. \tag{23}$$

Hence, the summation for $y \leq y^*$ is bounded as

$$\sum_{y=0}^{y^*} \left(\frac{\varepsilon_n}{1 + \varepsilon_n} \right)^y q(y) \log\big(q(y)\big) \geq \sum_{y=0}^{y^*} \left(\frac{\varepsilon_n}{1 + \varepsilon_n} \right)^y (y + 1) e^{-\alpha} \big(\log(y + 1) - \alpha \big). \tag{24}$$

And, multiplying by the exponential factor $e^{\frac{\varepsilon_s x}{\varepsilon_n}}$, we have

$$\sum_{y=0}^{y^*} e^{-\varepsilon_s x}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1)\log(y+1) - \sum_{y=0}^{y^*} e^{-\varepsilon_s x}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1)\alpha. \tag{25}$$

Both summands vanish as $y \to \infty$. The second has the form

$$e^{-\varepsilon_s x}\sum_{y=0}^{y^*}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1)\alpha, \tag{26}$$

which decays exponentially in x, since the sum satisfies

$$\sum_{y=0}^{y^*}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1) \leq \sum_{y=0}^{\infty}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1) = (1+\varepsilon_n)^2, \tag{27}$$

and $e^{-\varepsilon_s x}\alpha(1+\varepsilon_n)^2$ vanishes for large x. Similarly, the summation

$$\sum_{y=0}^{y^*}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y (y+1)\log(y+1) \tag{28}$$

remains bounded, since it is the partial sum of a convergent series, with n-th coefficient $\beta^n(n+1)\log(n+1)$ and $\beta = \varepsilon_n/(1+\varepsilon_n) < 1$. This is verified by the checking the ratio test, as

$$\lim_{n\to\infty}\frac{\beta^n(n+1)\log(n+1)}{\beta^{n-1}n\log n} = \beta < 1. \tag{29}$$

Boundedness of the partial sum implies that, after multiplying times an exponential factor $e^{-\varepsilon_s x}$, the first summand vanishes as $x \to \infty$.

Next, we consider the remainder of the summation in Eq. (22),

$$\sum_{y=y^*+1}^{\infty}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y q(y)\log\big(q(y)\big). \tag{30}$$

Clearly, $q(0) = e^{-\alpha} \leq q(y) \leq 1$ and therefore $-\alpha \leq \log q(y) \leq 0$, so each summand is negative and bounded by

$$\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y q(y)\log\big(q(y)\big) \geq -\alpha\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y q(y) \geq -\alpha\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y. \tag{31}$$

Summing over y,

$$\sum_{y=y^*+1}^{\infty}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y q(y)\log\big(q(y)\big) \geq -\alpha(1+\varepsilon_n)\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^{y^*+1}. \tag{32}$$

Using that $\alpha = \varepsilon_s x\big(1+\frac{1}{\varepsilon_n}\big)$ and taking into account the denominator $\varepsilon_s x(1+\varepsilon_n)$ in Eq. (22), we must study the behaviour of

$$-\frac{\varepsilon_n+1}{\varepsilon_n}e^{\frac{\varepsilon_s x}{\varepsilon_n}}\left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^{y^*+1} = -\frac{\varepsilon_n+1}{\varepsilon_n}e^{\frac{\varepsilon_s x}{\varepsilon_n}-(y^*+1)\log\left(1+\frac{1}{\varepsilon_n}\right)} \tag{33}$$

as $x \to \infty$. By construction, $y^* + 1 > \alpha$, and therefore

$$\frac{\varepsilon_s x}{\varepsilon_n} - (y^*+1)\log\left(1+\frac{1}{\varepsilon_n}\right) < \frac{\varepsilon_s x}{\varepsilon_n} - \varepsilon_s x\left(1+\frac{1}{\varepsilon_n}\right)\log\left(1+\frac{1}{\varepsilon_n}\right) \tag{34}$$

$$= \varepsilon_s x\left(\frac{1}{\varepsilon_n} - \left(1+\frac{1}{\varepsilon_n}\right)\log\left(1+\frac{1}{\varepsilon_n}\right)\right). \tag{35}$$

Since $t \leq (1+t)\log(1+t)$ for $t > 0$, a fact which follows from the inequality $\log(1+t) \leq t$, the left-hand side of Eq. (34) is strictly upper bounded by a function ax, with $a < 0$. Hence, the function in Eq. (33) vanishes exponentially as $x \to \infty$, and so does the term

$$\frac{e^{\frac{\varepsilon_s x}{\varepsilon_n}}}{\varepsilon_s x (1 + \varepsilon_n)} \sum_{y=y^*+1}^{\infty} \left(\frac{\varepsilon_n}{1+\varepsilon_n}\right)^y q(y) \log(q(y)). \tag{36}$$

This proves the limit in Eq. (22) and that Eq. (21) holds with equality.

APPENDIX B
ASYMPTOTICS AT LOW ε_s FOR $Z = 0$

The mutual information is given by

$$I(X;Y) = -\sum_x P(x) \sum_{y=0}^{\infty} Q(y|x) \log\left(\sum_{x' \in \mathcal{X}} P(x') e^{\varepsilon_s (x-x')} \left(\frac{x'}{x}\right)^y\right). \tag{37}$$

Using the Taylor expansion of the exponential $e^t = 1 + t + \frac{1}{2}t^2 + \mathrm{O}(t^3)$, we notice that there are only three possible channel outputs to order ε_s^3, namely

$$y = 0, \quad Q(y|x) = 1 - \varepsilon_s x + \frac{1}{2}\varepsilon_s^2 x^2 + \mathrm{O}(\varepsilon_s^3) \tag{38}$$

$$y = 1, \quad Q(y|x) = \varepsilon_s x - \varepsilon_s^2 x^2 + \mathrm{O}(\varepsilon_s^3) \tag{39}$$

$$y = 2, \quad Q(y|x) = \frac{1}{2}\varepsilon_s^2 x^2 + \mathrm{O}(\varepsilon_s^3) \tag{40}$$

$$y > 2, \quad Q(y|x) = \mathrm{O}(\varepsilon_s^3). \tag{41}$$

Since each of these cases behaves differently, we examine them separately.

We rewrite the variable in the $\log(\cdot)$ in Eq. (37) with the appropriate approximation. When the output is $y = 0$, the variable is

$$\sum_{x' \in \mathcal{X}} P(x') e^{\varepsilon_s (x-x')} = \sum_{x' \in \mathcal{X}} P(x')\left(1 + \varepsilon_s (x-x') + \frac{1}{2}\varepsilon_s^2 (x-x')^2 + \mathrm{O}(\varepsilon_s^3)\right) \tag{42}$$

$$= 1 + \varepsilon_s (x - \mu_1) + \frac{1}{2}\varepsilon_s^2 (x^2 + \mu_2 - 2x\mu_1) + \mathrm{O}(\varepsilon_s^3). \tag{43}$$

Taking logarithms, and using the formula $\log(1+t) = t - \frac{1}{2}t^2 + \mathrm{O}(t^3)$, we obtain

$$\varepsilon_s (x - \mu_1) + \frac{1}{2}\varepsilon_s^2 (x^2 + \mu_2 - 2x\mu_1) - \frac{1}{2}\varepsilon_s^2 (x^2 + \mu_1^2 - 2x\mu_1) + \mathrm{O}(\varepsilon_s^3) \tag{44}$$

$$= \varepsilon_s (x - \mu_1) + \frac{1}{2}\varepsilon_s^2 \left(\mu_2 - \mu_1^2\right) + \mathrm{O}(\varepsilon_s^3). \tag{45}$$

For $y = 1$, the variable in the logarithm in Eq. (37) is

$$\sum_{x' \in \mathcal{X}} P(x') e^{\varepsilon_s (x-x')} \frac{x'}{x} = \frac{1}{x} \sum_{x' \in \mathcal{X}} P(x') x' \left(1 + \varepsilon_s (x-x') + \mathrm{O}(\varepsilon_s^2)\right) \tag{46}$$

$$= \frac{\mu_1}{x} \left(1 + \frac{\varepsilon_s}{\mu_1} (x\mu_1 - \mu_2) + \mathrm{O}(\varepsilon_s^2)\right). \tag{47}$$

Taking logarithms, and using the Taylor expansion, we get

$$\log \frac{\mu_1}{x} + \frac{\varepsilon_s}{\mu_1}\big(x\mu_1 - \mu_2\big) + \mathrm{O}(\varepsilon_s^2). \tag{48}$$

We will later verify that no higher-order terms are required.

At last, for $y = 2$, the variable in the logarithm in Eq. (37) is

$$\sum_{x' \in \mathcal{X}} P(x')e^{\varepsilon_s(x-x')}\frac{x'^2}{x^2} = \frac{1}{x^2}\sum_{x' \in \mathcal{X}} P(x')x'^2\Big(1 + \mathrm{O}(\varepsilon_s)\Big) \tag{49}$$

$$= \frac{\mu_2}{x^2}\big(1 + \mathrm{O}(\varepsilon_s)\big). \tag{50}$$

Again, taking logarithms, and using the Taylor expansion, we get

$$\log \frac{\mu_2}{x^2} + \mathrm{O}(\varepsilon_s). \tag{51}$$

Later, we will verify that no higher-order terms are required.

After carrying out the averaging over y, we first combine Eqs. (45), (48) and (51) with the probabilities in Eqs. (38)–(40) and then group all terms up to $\mathrm{O}(\varepsilon_s^3)$ to derive

$$\Big(1 - \varepsilon_s x + \tfrac{1}{2}\varepsilon_s^2 x^2\Big)\Big(\varepsilon_s(x - \mu_1) + \tfrac{1}{2}\varepsilon_s^2\big(\mu_2 - \mu_1^2\big)\Big) +$$

$$+ \big(\varepsilon_s x - \varepsilon_s^2 x^2\big)\Big(\log \frac{\mu_1}{x} + \frac{\varepsilon_s}{\mu_1}\big(x\mu_1 - \mu_2\big)\Big) + \tfrac{1}{2}\varepsilon_s^2 x^2 \log \frac{\mu_2}{x^2} + \mathrm{O}(\varepsilon_s^3) \tag{52}$$

$$= \varepsilon_s(x - \mu_1) + \tfrac{1}{2}\varepsilon_s^2\big(\mu_2 - \mu_1^2\big) - \varepsilon_s^2(x^2 - x\mu_1) + \varepsilon_s x \log \frac{\mu_1}{x}$$

$$- \varepsilon_s^2 x^2 \log \frac{\mu_1}{x} + \varepsilon_s^2 \frac{1}{\mu_1}\big(x^2\mu_1 - x\mu_2\big) + \tfrac{1}{2}\varepsilon_s^2 x^2 \log \frac{\mu_2}{x^2} + \mathrm{O}(\varepsilon_s^3). \tag{53}$$

The expectation over x is straightforward, and gives the desired $I(X; Y)$.

APPENDIX C

ASYMPTOTICS AT LOW ε_s FOR POISSON NOISE

The mutual information is given by

$$I(X;Y) = -\sum_x P(x) \sum_{y=0}^{\infty} Q(y|x) \log \left(\sum_{x' \in \mathcal{X}} P(x')e^{\varepsilon_s(x-x')} \frac{\Big(1 + \frac{\varepsilon_s x'}{\varepsilon_n}\Big)^y}{\Big(1 + \frac{\varepsilon_s x}{\varepsilon_n}\Big)^y} \right). \tag{54}$$

Using the Taylor expansion of the exponential, and neglecting terms of order higher than ε_s^3, the channel output law is given by

$$Q(y|x) = \frac{\varepsilon_n^y}{y!}e^{-\varepsilon_n}\left(1 + \varepsilon_s x\left(\frac{y}{\varepsilon_n} - 1\right) + \frac{(\varepsilon_s x)^2}{2}\left(\frac{y(y-1)}{\varepsilon_n^2} + 1 - \frac{2y}{\varepsilon_n}\right)\right) + \mathrm{O}(\varepsilon_s^3). \tag{55}$$

We next examine the logarithm in Eq. (54). First, the Taylor expansions of $(1+t)^y$ and $(1+t)^{-y}$ yield

$$\frac{\Big(1 + \frac{\varepsilon_s x'}{\varepsilon_n}\Big)^y}{\Big(1 + \frac{\varepsilon_s x}{\varepsilon_n}\Big)^y} = 1 + \frac{\varepsilon_s}{\varepsilon_n}y(x' - x) + \frac{\varepsilon_s^2}{2\varepsilon_n^2}\big(y^2(x'-x)^2 - y(x'^2 - x^2)\big) + \mathrm{O}(\varepsilon_s^3). \tag{56}$$

Similarly, using the expansion of the exponential, we have

$$e^{\varepsilon_s(x-x')}\frac{\left(1+\frac{\varepsilon_s x'}{\varepsilon_n}\right)^y}{\left(1+\frac{\varepsilon_s x}{\varepsilon_n}\right)^y} = 1 + \varepsilon_s\left(\frac{y}{\varepsilon_n}-1\right)(x'-x) + \frac{\varepsilon_s^2}{2}\left(\left(1-\frac{y}{\varepsilon_n}\right)^2(x-x')^2 + \frac{y(x^2-x'^2)}{\varepsilon_n^2}\right) + \mathrm{O}(\varepsilon_s^3).$$

(57)

Now, carrying out the expectation over x' we obtain

$$\sum_{x'\in\mathcal{X}} P(x')e^{\varepsilon_s(x-x')}\frac{\left(1+\frac{\varepsilon_s x'}{\varepsilon_n}\right)^y}{\left(1+\frac{\varepsilon_s x}{\varepsilon_n}\right)^y} = 1 + \varepsilon_s\left(\frac{y}{\varepsilon_n}-1\right)(\mu_1-x) + \frac{\varepsilon_s^2}{2}\left(\left(1-\frac{y}{\varepsilon_n}\right)^2(x^2+\mu_2-2x\mu_1) + \frac{y(x^2-\mu_2)}{\varepsilon_n^2}\right) + \mathrm{O}(\varepsilon_s^3).$$

(58)

Next, using the expansion of the logarithm, we obtain

$$\log\left(\sum_{x'\in\mathcal{X}} P(x')e^{\varepsilon_s(x-x')}\frac{\left(1+\frac{\varepsilon_s x'}{\varepsilon_n}\right)^y}{\left(1+\frac{\varepsilon_s x}{\varepsilon_n}\right)^y}\right) = \varepsilon_s\left(\frac{y}{\varepsilon_n}-1\right)(\mu_1-x) + \frac{\varepsilon_s^2}{2}\left(\left(1-\frac{y}{\varepsilon_n}\right)^2(\mu_2-\mu_1^2) + \frac{y(x^2-\mu_2)}{\varepsilon_n^2}\right) + \mathrm{O}(\varepsilon_s^3).$$

(59)

Now, multiplying by the channel law, we get for given x and y

$$\frac{\varepsilon_n^y}{y!}e^{-\varepsilon_n}\left(\varepsilon_s\left(\frac{y}{\varepsilon_n}-1\right)(\mu_1-x) + \frac{\varepsilon_s^2}{2}\left(\left(1-\frac{y}{\varepsilon_n}\right)^2(\mu_2-\mu_1^2+2\mu_1 x-2x^2) + \frac{y(x^2-\mu_2)}{\varepsilon_n^2}\right) + \mathrm{O}(\varepsilon_s^3)\right).$$ (60)

After carrying out the expectation over x, some terms cancel to give

$$\frac{\varepsilon_n^y}{y!}e^{-\varepsilon_n}\left(\frac{\varepsilon_s^2}{2}\left(1-\frac{y}{\varepsilon_n}\right)^2(\mu_1^2-\mu_2) + \mathrm{O}(\varepsilon_s^3)\right).$$

(61)

As a final step, we sum over y to obtain the mutual information,

$$\sum_{y=0}^{\infty}\frac{\varepsilon_n^y}{y!}e^{-\varepsilon_n}\left(\frac{\varepsilon_s^2}{2}\left(1-\frac{2y}{\varepsilon_n}+\frac{y^2}{\varepsilon_n^2}\right)(\mu_1^2-\mu_2) + \mathrm{O}(\varepsilon_s^3)\right) = \frac{\varepsilon_s^2}{2\varepsilon_n}(\mu_1^2-\mu_2) + \mathrm{O}(\varepsilon_s^3).$$

(62)

APPENDIX D

ASYMPTOTICS AT LOW ε_s FOR GEOMETRIC NOISE

The mutual information is given by

$$I(X;Y) = -\sum_x P(x)\sum_{y=0}^{\infty} Q(y|x)\log\left(\frac{\sum_{x'\in\mathcal{X}} P(x')Q(y|x')}{Q(y|x)}\right),$$

(63)

where $Q(y|x)$ is given by Eq. (3).

As it happened in the discrete-time Poisson channel, the Taylor expansion of the exponential implies that there are only three possible channel outputs s to order ε_s^3, that is,

$$s=0, \quad P(s|x) = 1 - \varepsilon_s x + \frac{1}{2}(\varepsilon_s x)^2 + \mathrm{O}(\varepsilon_s^3)$$ (64)

$$s=1, \quad P(s|x) = \varepsilon_s x - (\varepsilon_s x)^2 + \mathrm{O}(\varepsilon_s^3)$$ (65)

$$s=2, \quad P(s|x) = \frac{1}{2}(\varepsilon_s x)^2 + \mathrm{O}(\varepsilon_s^3)$$ (66)

$$s>2, \quad P(s|x) = \mathrm{O}(\varepsilon_s^3).$$ (67)

Hence the channel output $y = s + z$ only includes these contributions. We distinguish three cases, viz. $y = 0$, $y = 1$, and $y \geq 2$.

In the first case, $y = s = z = 0$, and $Q(y|x)$ becomes

$$Q(y|x) = \frac{1}{1 + \varepsilon_n}\left(1 - \varepsilon_s x + \frac{1}{2}(\varepsilon_s x)^2 + \mathrm{O}(\varepsilon_s^3)\right). \tag{68}$$

For $y = 1$, we combine the outputs $s = 0$ and $s = 1$,

$$Q(y|x) = \frac{\varepsilon_n}{(1 + \varepsilon_n)^2}\left(1 + \frac{\varepsilon_s}{\varepsilon_n}x - \varepsilon_s^2 x^2\left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right) + \mathrm{O}(\varepsilon_s^3)\right). \tag{69}$$

For $y \geq 2$, we combine the outputs $s = 0$, $s = 1$, and $s = 2$,

$$Q(y|x) = \frac{\varepsilon_n^y}{(1 + \varepsilon_n)^{y+1}}\left(1 + \frac{\varepsilon_s}{\varepsilon_n}x + \frac{\varepsilon_s^2}{2\varepsilon_n^2}x^2 + \mathrm{O}(\varepsilon_s^3)\right), \tag{70}$$

after combining some terms together.

We next rewrite the numerator and denominator in the $\log(\cdot)$ in Eq. (63) with the appropriate approximation. For $y = 0$, the common term $(1 + \varepsilon_n)^{-1}$ cancels, and the numerator is

$$\sum_{x' \in \mathcal{X}} Q(y|x')P(x') = \left(1 - \varepsilon_s \mu_1 + \frac{1}{2}\varepsilon_s^2 \mu_2\right) + \mathrm{O}(\varepsilon_s^3). \tag{71}$$

In the denominator, we keep the expansion

$$1 - \varepsilon_s x + \tfrac{1}{2}(\varepsilon_s x)^2\, \mathrm{O}(\varepsilon_s^3). \tag{72}$$

Taking logarithms of Eqs. (71) and (72), using a Taylor expansion, and combining numerator and denominator, we obtain

$$\log\left(1 - \varepsilon_s \mu_1 + \frac{1}{2}\varepsilon_s^2 \mu_2 + \mathrm{O}(\varepsilon_s^3)\right) - \log\left(1 - \varepsilon_s x + \frac{1}{2}(\varepsilon_s x)^2 + \mathrm{O}(\varepsilon_s^3)\right) \tag{73}$$

$$= -\varepsilon_s(\mu_1 - x) + \frac{1}{2}\varepsilon_s^2(\mu_2 - \mu_1^2) + \mathrm{O}(\varepsilon_s^3). \tag{74}$$

For $y = 1$, we use Eq. (69). Summing over x' in the numerator, we get

$$\sum_{x' \in \mathcal{X}} Q(y|x')P(x') = \left(1 + \frac{\varepsilon_s}{\varepsilon_n}\mu_1 - \left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)\mu_2 \varepsilon_s^2 + \mathrm{O}(\varepsilon_s^3)\right), \tag{75}$$

with the agreement that a common term $\varepsilon_n/(1 + \varepsilon_n)^2$ has been cancelled.

Combining numerator and denominator, taking logarithms, and using the Taylor expansion of the logarithm, we obtain

$$\frac{\varepsilon_s}{\varepsilon_n}\mu_1 - \left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)\mu_2 \varepsilon_s^2 - \frac{\varepsilon_s^2}{2\varepsilon_n^2}\mu_1^2 - \frac{\varepsilon_s}{\varepsilon_n}x + \left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)x^2 \varepsilon_s^2 + \frac{\varepsilon_s^2}{2\varepsilon_n^2}x^2 + \mathrm{O}(\varepsilon_s^3) \tag{76}$$

$$= \frac{(\mu_1 - x)\varepsilon_s}{\varepsilon_n} - \left(\left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)\mu_2 + \frac{\mu_1^2}{2\varepsilon_n^2} - \frac{x^2(1 + \varepsilon_n)^2}{2\varepsilon_n^2}\right)\varepsilon_s^2 + \mathrm{O}(\varepsilon_s^3). \tag{77}$$

If the output is $y \geq 2$, in an analogous way we use Eq. (70) to rewrite the logarithm of the ratio of numerator and denominator as

$$\log\left(1 + \frac{\varepsilon_s}{\varepsilon_n}\mu_1 + \frac{\varepsilon_s^2}{2\varepsilon_n^2}\mu_2 + \mathrm{O}(\varepsilon_s^3)\right) - \log\left(1 + \frac{\varepsilon_s}{\varepsilon_n}x + \frac{\varepsilon_s^2}{2\varepsilon_n^2}x^2 + \mathrm{O}(\varepsilon_s^3)\right). \tag{78}$$

Using now the Taylor expansion of the logarithm, we obtain

$$\frac{\varepsilon_s}{\varepsilon_n}(\mu_1 - x) + \frac{\varepsilon_s^2}{2\varepsilon_n^2}(\mu_2 - \mu_1^2) + \mathrm{O}(\varepsilon_s^3). \tag{79}$$

The remaining steps are the averaging over x and y. We first carry out the expectation over x. From Eq. (74), the averaging over x yields

$$\sum_x \frac{1}{1 + \varepsilon_n}\left(1 - \varepsilon_s x + \frac{1}{2}(\varepsilon_s x)^2\right)\left((x - \mu_1)\varepsilon_s + \frac{1}{2}\varepsilon_s^2(\mu_2 - \mu_1^2)\right) + \mathrm{O}(\varepsilon_s^3) \tag{80}$$

$$= \frac{1}{1 + \varepsilon_n}\frac{1}{2}\varepsilon_s^2(\mu_1^2 - \mu_2) + \mathrm{O}(\varepsilon_s^3). \tag{81}$$

Similarly, from Eq. (77) we obtain (bar for a constant factor $\frac{\varepsilon_n}{(1+\varepsilon_n)^2}$)

$$\sum_x \left(1 + \frac{\varepsilon_s}{\varepsilon_n}x - \varepsilon_s^2 x^2\left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)\right) \times$$

$$\times \left(\frac{(\mu_1 - x)\varepsilon_s}{\varepsilon_n} - \left(\left(\frac{1}{2} + \frac{1}{\varepsilon_n}\right)\mu_2 + \frac{\mu_1^2}{2\varepsilon_n^2} - \frac{x^2(1 + \varepsilon_n)^2}{2\varepsilon_n^2}\right)\varepsilon_s^2\right) + \mathrm{O}(\varepsilon_s^3) \tag{82}$$

$$= (\mu_1^2 - \mu_2)\frac{\varepsilon_s^2}{2\varepsilon_n^2} + \mathrm{O}(\varepsilon_s^3). \tag{83}$$

And finally, from Eq. (79), for $y \geq 2$, we get

$$\sum_x \frac{\varepsilon_n^y}{(1 + \varepsilon_n)^{y+1}}\left(1 + \frac{\varepsilon_s}{\varepsilon_n}x + \frac{\varepsilon_s^2}{2\varepsilon_n^2}x^2\right)\left(\frac{\varepsilon_s}{\varepsilon_n}(\mu_1 - x) + \frac{\varepsilon_s^2}{2\varepsilon_n^2}(\mu_2 - \mu_1^2)\right) + \mathrm{O}(\varepsilon_s^3) \tag{84}$$

$$= \frac{\varepsilon_n^y}{(1 + \varepsilon_n)^{y+1}}\frac{\varepsilon_s^2}{2\varepsilon_n^2}(\mu_1^2 - \mu_2) + \mathrm{O}(\varepsilon_s^3). \tag{85}$$

The summation over $y \geq 2$ can be carried out and yields

$$\sum_{y=2}^{\infty} \frac{\varepsilon_n^y}{(1 + \varepsilon_n)^y} = \frac{\varepsilon_n^2}{1 + \varepsilon_n}. \tag{86}$$

Then, combining Eq. (86) into Eq. (85), and summing with Eqs. (81) and (83) (including the factor $\frac{\varepsilon_n}{(1+\varepsilon_n)^2}$), we obtain

$$\frac{1}{1 + \varepsilon_n}\frac{1}{2}\varepsilon_s^2(\mu_1^2 - \mu_2)\left(1 + \frac{1}{\varepsilon_n(1 + \varepsilon_n)} + \frac{1}{\varepsilon_n^2}\frac{\varepsilon_n^2}{1 + \varepsilon_n}\right) + \mathrm{O}(\varepsilon_s^3) \tag{87}$$

$$= \frac{1}{2}\varepsilon_s^2(\mu_1^2 - \mu_2)\frac{1}{\varepsilon_n} + \mathrm{O}(\varepsilon_s^3). \tag{88}$$

The expansion for $I(X;Y)$ follows.

REFERENCES

[1] V. Prelov and S. Verdú, "Second order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, August 2004.

[2] A. Martinez, "Information rates of radiation as a photon gas," *Phys. Rev. A*, vol. 77, no. 3, pp. 032 116–1/7, 2008.

[3] ——, "Spectral efficiency of optical direct detection," *J. Opt. Soc. Am. B*, vol. 24, no. 4, pp. 739–749, April 2007.

[4] A. Lapidoth and S. M. Moser, "Bounds on the capacity of the discrete-time Poisson channel," in *Proceedings of the 41st Allerton Conf. on Communication, Control, and Computing*, October 2003.

[5] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, September 1990.

[6] ——, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.
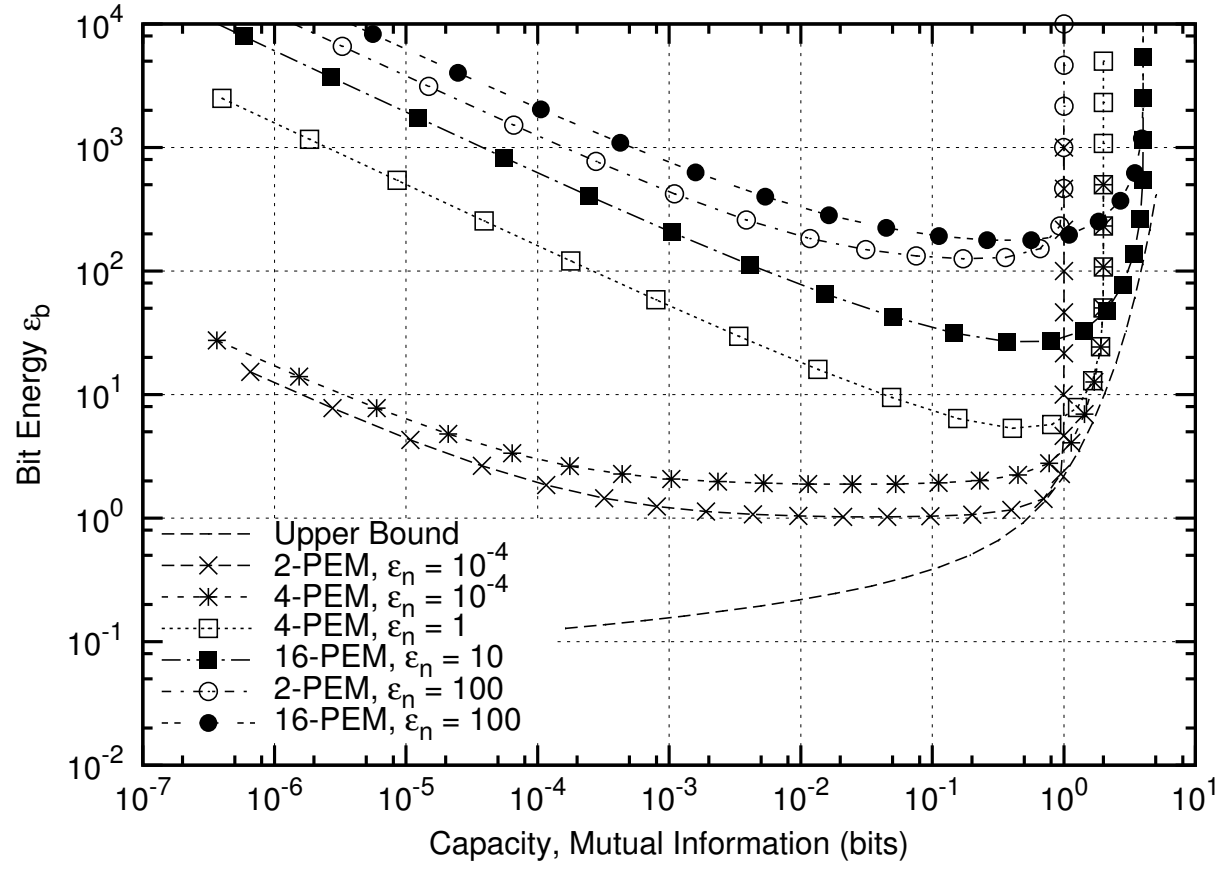
Fig. 1.   Bit energy ε_b as a function of the mutual information for uniform 2^m-PEM and varying ε_n (geometric noise).
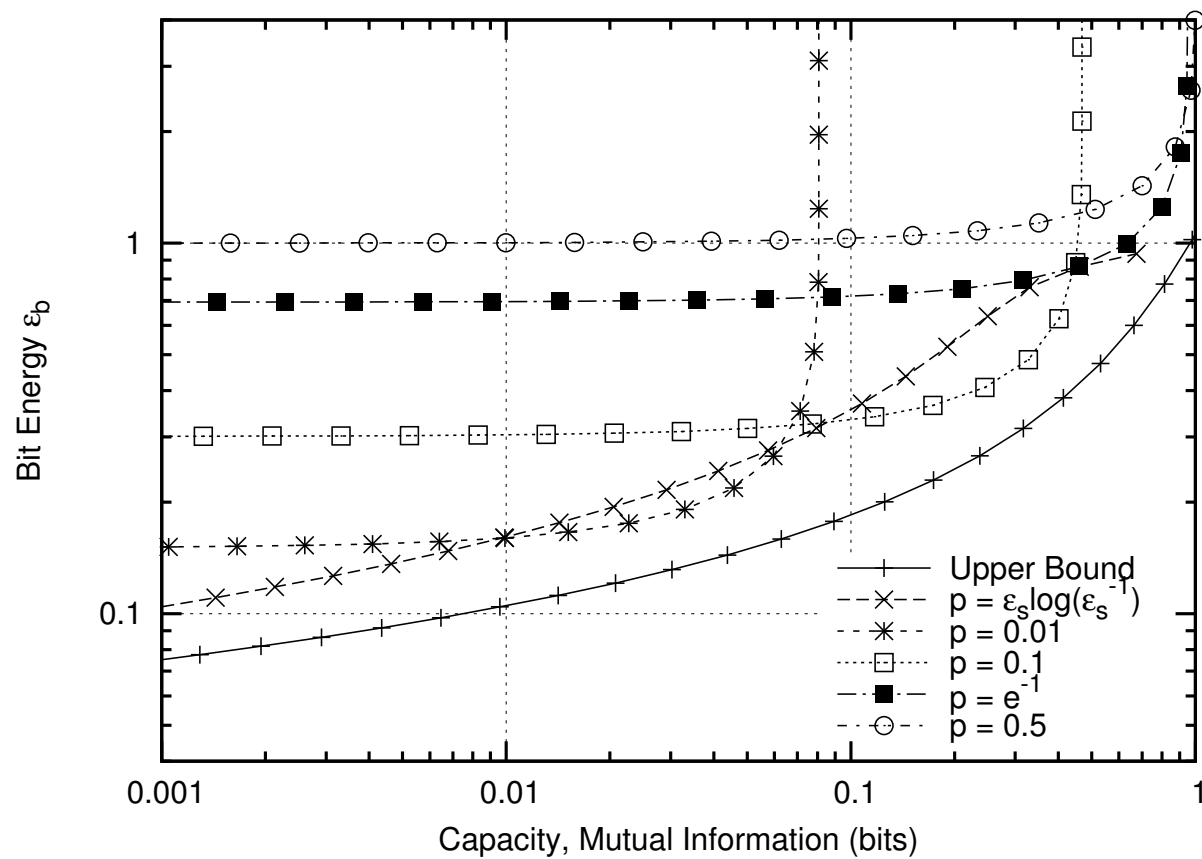
Fig. 2.   Bit energy ε_b as a function of the mutual information or the capacity (noiseless channel, $Z = 0$).